MULTIMEDIA UNIVERSITY

STUDENT ID NO

MULTIMEDIA UNIVERSITY

FINAL EXAMINATION

TRIMESTER 2, 2019/2020

DSP5614 – PROBABILITY AND STATISTICS

(*for diploma students only*)

2 MARCH 2020

9.00 a.m. – 11.00 a.m.
(2 Hours)

INSTRUCTIONS TO STUDENT

1. This question paper consists of **3** pages excluding cover page and appendix.

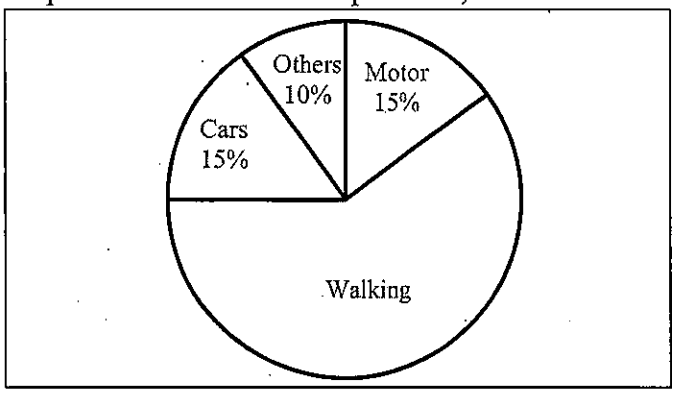
2. Attempt **ALL FOUR (4)** questions. The marks distribution for each question is given.

3. Write your answers in the answer booklet provided.

4. Key formulae are given in the Appendix.

Question 1

a) Given a pie chart of transportation method to campus for 1,500 students in MMU.



Construct a distribution table consisting type of transportation, frequency and relative frequency to represent the data above. (5 marks)

b) Given a data set for the number of people living in a house: 8, 11, 14, 7, 6, 5, 5.

i) Identify the mean and median for the data set. (4 marks)

ii) Based on your answer in (i), what is the shape of the distribution? (2 marks)

c) Given a data set of time taken by 30 random beach goer to swim 50m.

65	73	59	58	33	70	56	71	55	73
48	52	67	46	75	51	75	55	32	54
76	78	34	62	68	79	39	72	64	46

i) Construct a grouped data table starting with $30 \leqslant 40$. (4 marks)

Based on the grouped data table in (i),

ii) state the mode class. (1 mark)

iii) calculate the mean and standard deviation. (7 marks)

d) Given a stem-and-leaf diagram of marks for 14 students who failed in their statistic quiz.

Stem	Leaves
1	7
2	0 1 4 9
3	4 6 8
4	4 4 5 6 7 8

Key: 1 | 7 = 17

i) Find the quartiles. (6 marks)

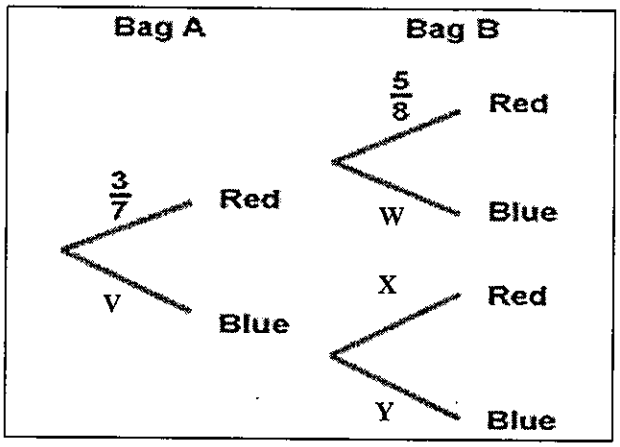
ii) List out the 5 number summary. (1 mark)

[TOTAL: 30 MARKS]

Continued...

Question 2

a) Bag A contains 6 red balls and 8 blue balls. Bag B contains 5 red balls and 3 blue balls. A ball is taken from each bag in turn and the following tree diagram shows the possible outcome.



i) Complete the probabilities of V, W, X and Y on the tree diagram. (4 marks)

ii) Calculate the probability of getting at least one red ball. (4 marks)

b) A box contains 3 apples, 12 oranges and x bananas. When a fruit is chosen at random from the box, the probability of getting an orange is 75%. Find the value of x. (4 marks)

c) The contingency table below shows a random sample of 100 runners and the terrain they prefer to run on.

Gender	Beach (B)	Hill (H)	Lake (L)	Total
Male (M)	12	17	16	45
Female (F)	13	17	x	55
Total	25	34	y	100

Find the value of

i) x and y. (2 marks)

ii) $P(F \cup L)$. (3 marks)

iii) $P(B \mid M)$. (3 marks)

[TOTAL: 20 MARKS]

Question 3

a) A normal chili plant has a mean length of 8.3cm and standard deviation of 2.4cm. A biotechnologist claimed that longer chili can be produced by using nanotechnology. To check on this, a random sample of 36 chilies was taken and gave a sample mean length of 9.1cm. At 1% significance value, can we accept the claim made by the biotechnologist? (10 marks)

Continued...

b) A previous survey conducted by Student Representative Council (SRC) to see the satisfaction level of students with the service provided by the café operator. The result showed that less than half of the students are satisfied with the service provided. To gain more information on the latest situation, SRC has done another survey and the result is shown as below:

NO	NO	YES	YES	NO	NO	YES	YES	NO	YES
YES	YES	NO	NO	NO	YES	NO	NO	YES	NO
NO	NO	NO	YES	YES	NO	YES	NO	NO	NO
YES	NO	NO	YES	NO	YES	NO	NO	YES	YES

Based on the result above, can we conclude that the previous study is true at 10% significance level? (12 marks)

c) A car magazine is comparing the average expenses per day for two luxury car due to the popular opinion that Model V has lower expenses per day compared to Model W. A random sample of 25 Model V and 35 Model W gives a test statistics value of $z_0 = -1.75$ and p-value 0.0401. At 5% significance level, what can be concluded from the comparison? (3 marks)

[TOTAL: 25 MARKS]

Question 4

a) A final year student in Diploma of Fashion and Design would like to study the relationship between shoe size and height. He recorded the details of 10 persons as follow.

Shoe size, x	Height (cm), y
4.5	164.5
10.5	170
6.5	166
9.5	170
9	168
7.5	167.5
8.5	166.5
12	171
9.5	169
5	166

i) Calculate SS_{xx}, SS_{yy} and SS_{xy} based on the information given. (9 marks)

ii) Calculate the correlation coefficient, r and interpret the answer. (4 marks)

iii) Find the linear regression equation $\hat{y} = \beta_0 + \beta_1 x$. (5 marks)

iv) Predict the height of a person if the shoe size is 8. (2 marks)

v) Compute the value of *SST*, *SSR* and *SSE*. (3 marks)

vi) Calculate the standard error of the estimate, s_e. (2 marks)

[TOTAL: 25 MARKS]

End of paper.

APPENDIX – KEY FORMULA

	Ungrouped Data	Grouped Data
Mean	$\bar{x} = \frac{\sum x_i}{n}$ where n : sample size	$\bar{x} = \frac{\sum m_i f_i}{\sum f_i}$ where m : class midpoint, f : class frequency
Variance	$s^2 = \frac{1}{n-1}\left[\sum x_i^{\,2} - \frac{\left(\sum x_i\right)^2}{n}\right]$ where n : sample size	$s^2 = \frac{1}{\left(\sum f_i\right)-1}\left[\sum m_i^{\,2} f_i - \frac{\left(\sum m_i f_i\right)^2}{\sum f_i}\right]$ where m : class midpoint, f : class frequency

NORMAL AND STANDARD NORMAL PROBABILITY DISTRIBUTION

- z-value (observed value) for an x value : $Z = \frac{x-\mu}{\sigma}$

HYPOTHESIS TESTING

Population	Mean	Proportion
1	$Z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$	$Z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ where $\hat{p} = \frac{x}{n}$
2	$Z = \frac{\left(\bar{x}_1-\bar{x}_2\right)-\left(\mu_1-\mu_2\right)}{\sqrt{\frac{\sigma_1^{\,2}}{n_1}+\frac{\sigma_2^{\,2}}{n_2}}}$	$Z = \frac{\left(\hat{p}_1-\hat{p}_2\right)-\left(p_1-p_2\right)}{\sqrt{\hat{p}_p\left(1-\hat{p}_p\right)}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ where $\hat{p}_p = \frac{x_1+x_2}{n_1+n_2}$

Continued...

LINEAR REGRESSION ANALYSIS

- **Least square regression equation,** $\hat{y} = \beta_0 + \beta_1 x$ where $\beta_1 = \dfrac{S_{xy}}{S_{xx}}$ and $\beta_0 = \bar{y} - \beta_1 \bar{x}$

- **Correlation coefficient,** $r = \dfrac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$ where

$S_{xx} = \sum x^2 - \dfrac{(\sum x)^2}{n}$	$S_{yy} = \sum y^2 - \dfrac{(\sum y)^2}{n}$	$S_{xy} = \sum xy - \dfrac{(\sum x)(\sum y)}{n}$

- **Sum of Square**

$SST = SSR + SSE$	$SST = S_{yy}$	$SSR = \dfrac{(S_{xy})^2}{S_{xx}}$

- **Test of Significance for Regression Slope**

Standard Error, $s_e = \sqrt{\dfrac{SSE}{n-2}}$	Regression t-Test $t = \dfrac{\beta_1}{s_e / \sqrt{Sxx}}$	Correlation t-Test $t = r \cdot \sqrt{\dfrac{n-2}{1-r^2}}$

End of page.